ADDRESSING DATA SHORTCOMINGS IN URBAN TRAFFIC MONITORING: HCMCTRAFFICDATASET - A NOVEL DATASET AND BASELINE MODEL

Tuyen Nguyen Quoc-Khanh⋆ *Minh-Son Dao*† *Thanh-Binh Nguyen*⋆

⋆ University of Science, Vietnam National University in Ho Chi Minh City
†National Institute of Information and Communications Technology (NICT), Japan

ABSTRACT

This paper addresses the challenges of Intelligent Transportation Systems (ITS) in developing countries, particularly with respect to motorcycle traffic, which is a dominant mode of transportation in many Southeast Asian countries. Existing traffic datasets may not adequately capture the diverse traffic conditions in these regions, where low-resolution cameras are commonly used for traffic monitoring. To address this gap, the authors present a novel dataset that captures traffic data in Ho Chi Minh City, Vietnam, with a focus on motorcycles. The paper includes a baseline model finetuned for motorcycle vehicle counting and a benchmarking analysis comparing the performance of different approaches in various traffic conditions and settings. This dataset and analysis can help facilitate the development of ITS solutions tailored to the unique traffic conditions of developing countries, with potential applications in the ASEAN Smart Cities Network and beyond.

***Index Terms*—** intelligent transportation systems, traffic datasets, vehicle counting, traffic monitoring

1. INTRODUCTION

Urbanization is a global phenomenon, with over half of the world's population now living in cities as of 2017 [1]. This trend is expected to continue, with urban areas projected to accommodate two-thirds of the global population in the near future [1]. As cities undergo rapid expansion, the strain on their infrastructure, notably transportation, becomes increasingly evident. To tackle this challenge, the concept of smart cities has emerged, utilizing digital solutions to enhance the efficiency of traditional networks and services [2]. An example of such an initiative is the ASEAN Smart Cities Network, which comprises 26 pilot cities across ASEAN member countries, aiming to foster smart and sustainable urban development [1]

Smart mobility, particularly leveraging Intelligent Transportation Systems (ITS), is crucial for smart cities, especially in regions with limited traffic monitoring resources, such as developing areas. However, implementing effective ITS faces several challenges, including vehicle detection and prediction in urban environments, particularly where low-resolution cameras are prevalent. Gathering precise data on current traffic conditions and regulating traffic flow are pivotal for an efficient ITS.

Existing research has delved into various traffic datasets, such as the STREETS dataset by Snyder et al. [3] and the PEMS-BAY and METR-LA dataset by Li et al. [4]. However, these datasets predominantly focus on four-wheeled vehicles, which may not accurately represent the diverse traffic conditions in developing countries, where motorcycles are often the primary mode of transportation [5]. Furthermore, these datasets often rely on high-quality sensors or cameras, which might not be readily available in all regions.

Therefore, there is a pressing need for datasets that capture the traffic dynamics of developing countries, where motorcycles play a significant role in urban transportation and where low-resolution cameras are commonly used for traffic monitoring. This paper addresses this gap by introducing a novel dataset that encompasses traffic data from urban areas of developing countries, with a specific emphasis on motorcycles. The dataset aims to facilitate research and development of ITS solutions tailored to the unique traffic conditions of developing countries, with potential applications in initiatives like the ASEAN Smart Cities Network and beyond.

In this academic paper, we aim to address the limitations of previous research by providing a new dataset that accurately represents the day-to-day traffic conditions of Ho Chi Minh City, Vietnam. This dataset is designed to increase the diversity of traffic conditions taken into account by researchers worldwide, thus addressing a gap in existing datasets that primarily focus on other regions. The contributions of our paper can be summarized as follows:

1. Novel traffic dataset release: We introduce a new dataset that is representative of the urban area of Ho Chi Minh City, Vietnam, providing valuable data on traffic patterns and dynamics in this specific context.
2. Baseline model finetuning: We finetune a baseline model specifically for motorcycle vehicle counting, which is a common mode of transportation in Ho Chi Minh City and other parts of Southeast Asia.

[1] https://asean.org/our-communities/asean-smart-cities-network/

3. Benchmarking of baseline methods: We compare the performance of our finetuned model with other popular methods on our dataset, as well as on several other widely-used datasets in the field. This benchmarking analysis provides insights into the effectiveness of different approaches in various traffic conditions and settings.

2. RELATED WORK

2.1. Other Datasets

Previous studies have examined traffic prediction datasets, focusing on collection methods and dataset characteristics. Traffic data is typically collected using traffic cameras or loop detectors, each with distinct advantages and limitations [6, 7]. Traffic cameras are cost-effective, flexible, and easy to maintain but can be affected by external factors like weather. Loop detectors offer higher accuracy for vehicle detection but may encounter reliability issues and require lane closures for maintenance [6]. For a detailed comparison between loop detectors and traffic cameras, refer to Ugo et al.'s study [7].

2.1.1. STREETS

The STREETS dataset, introduced in 2019, contains over 4 million images, including 3,000 annotated images collected over 2.5 months in Lake County, IL [3]. These images, sourced from various publicly available traffic cameras, provide a comprehensive view of traffic dynamics. The dataset covers diverse traffic and lighting conditions from both summer 2018 and summer 2019. Vehicle detection was performed using a custom fine-tuned Mask R-CNN model trained on 2,477 images, achieving a minimum detection confidence of 0.7. Additionally, the dataset was used to benchmark five baseline methods, demonstrating its utility.

2.1.2. METR-LA

Before STREETS, the primary benchmark for traffic prediction methods was the METR-LA dataset, available in the PyTorch Geometric Temporal dataset. It collects traffic data from loop detectors in Los Angeles County, featuring aggregated information recorded in 5-minute intervals over 4 months, from March 2012 to June 2012 [8]. This dataset's configuration was initially employed in introducing the Diffusion Convolutional Recurrent Neural Network (DCRNN) by Li et al. [4].

2.1.3. PEMS-Bay

PeMS-Bay, introduced by Li et al. in 2018, is a traffic dataset collected in the Bay Area using the Caltrans Performance Measurement System (PeMS)[4]. It covers six months from January 2017 to May 2017, featuring traffic speed data from 325 sensors aggregated into 5-minute intervals. Unlike STREETS and METR-LA, PeMS-Bay utilizes loop detectors instead of traffic cameras, providing more accurate measurements of average vehicle speed. It is also available in the PyTorch Geometric Temporal library.

2.1.4. Q-Traffic

The Q-Traffic dataset [9], includes three sub-datasets: the query sub-dataset, the traffic speed sub-dataset, and the road network dataset. The query sub-dataset, collected in Beijing, China from April 1st to May 31st, 2017, via the Baidu Map app, contains approximately 114 million user queries with timestamps, coordinates, and estimated travel times. The traffic speed sub-dataset, covering traffic speeds for each road segment within Beijing's congested 6th ring road, comprises around 266 million records smoothed with a 15-minute moving average. The road network dataset offers topology and geographical attributes for over 15,000 road segments, including width, direction, node IDs, and GPS coordinates, along with social attributes like weekdays, weekends, public holidays, peak hours, and off-peak hours.

2.1.5. UTD19

UTD19 is a vast traffic dataset collected from over 23,541 stationary loop detectors on urban roads across 40 global cities, making it the largest publicly available multi-city traffic dataset. Detailed information is available in the publication by Loder et al. [10] and on its website [11]. Accessible through a data repository[12], it offers data at 3-5 minute intervals, collected by the Institute for Transport Planning and Systems at ETH Zurich during a research campaign from 2017 to 2019. The dataset, geo-coded in WGS84 coordinates, has been preprocessed for consistent quality and is open to researchers worldwide for analysis.

2.1.6. IDD

IDD, proposed by Varma et al. in 2019, comprises over 10,000 annotated images collected from 182 drive sequences on Indian roads [13]. Unlike our dataset, IDD focuses on images captured by a vehicle driving in Bangalore, Hyderabad, and their outskirts, primarily for autonomous vehicle development. In contrast, our dataset's images are sourced from fixed-location traffic cameras.

2.1.7. WebCamT

WebCamT, by Zhang et al. [14], is a labeled real-world webcam traffic dataset with over 60 million images from 212 web cameras at a resolution of 352x240. Similar to ours, it shares traits like low resolution, frame rate, and high occlusion. However, while WebCamT mainly captures cars and other four-wheeled vehicles, our HCMCTrafficDataset focuses on motorcycles.

3. METHODOLOGY

In this section, we delve into the meticulous process of dataset collection and processing, offering insights into our methodology and the intricate graph structure underlying our dataset.

3.1. Data collection

3.1.1. Location

Ho Chi Minh City, Vietnam's largest city with nearly 9 million residents as of November 2019[15], faces frequent traffic congestion, especially during rush hours and in downtown areas, leading to frequent jams[16]. To address this issue and improve traffic management, the Ho Chi Minh City Department of Transportation has launched a publicly accessible website featuring live traffic cameras monitoring major road sections and intersections [2]. These stationary cameras provide continuous surveillance of specific road sections, intersections, or roundabouts 24/7, with varying specifications such as Field of View (FoV). The diversity in camera types and quality presents a realistic challenge for creating an Intelligent Transportation System with limited monitoring resources, particularly in developing countries with budget constraints for public infrastructure, including traffic monitoring. Unlike many prior datasets utilizing high-resolution cameras or loop detectors, our dataset reflects scenarios where such resources are unavailable due to factors like network bandwidth, costs, and security concerns, typical in urban areas within developing countries like Ho Chi Minh City.

3.1.2. Collection method

We collected data using traffic images from [3], updating every 5-10 seconds, focusing on a subset of 140 cameras. However, 13 cameras consistently provided unavailable images, potentially due to security concerns, including those capturing sensitive infrastructure like the US Embassy. Additionally, 5 cameras were not consistently accessible.

To balance data collection and network load, we cycled through the entire camera list with 60-second intervals between iterations. However, due to varying network configurations and VPN bandwidth, iteration durations fluctuated. To mitigate this, we provide an event log for each dataset day and a list of cameras with unavailable images, indicating sensitive areas restricting public access.

[2]http://giaothong.hochiminhcity.gov.vn/
[3]http://giaothong.hochiminhcity.gov.vn/

3.2. Data processing

In addition to providing traffic camera images, we offer traffic counts for each node detected from these images. We utilized the YOLOv5-XL model [17] as the baseline for detecting all vehicle types except motorcycles. For motorcycles, we finetuned a specific instance of the model to better suit Ho Chi Minh City's complex traffic and camera conditions, primarily focusing on underbones, the most common type of motorcycle in Vietnam.

To achieve this, we selected a random subset of 40 images per day within the dataset, labeled approximately 1000 images, including around 10% error images, as the golden subset, and then further finetuned the model with an additional subset of 100 images per day. We employed semi-supervised learning by using the model trained on the golden subset to label the additional images and applied hyperparameter tuning to determine the optimal configuration for motorcycle detection. Figure 1illustrates an example of data with annotations for each vehicle.

The resulting traffic counts are processed from the detection results, including filename, motorcycle count, vehicle count, and timestamp, determined by the time of the image request.

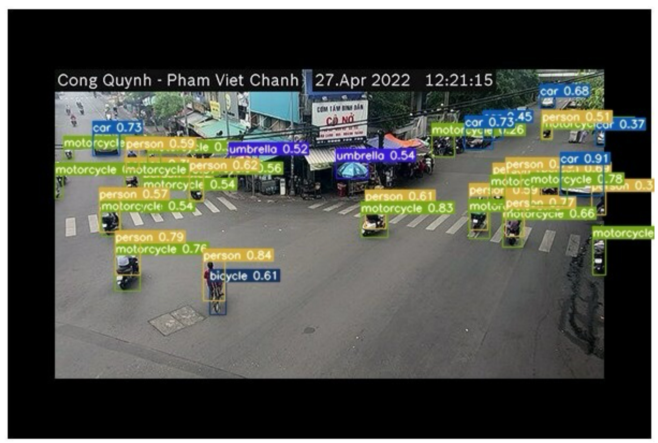


Fig. 1: An example of object detection using YOLOv5-XL, non-vehicle detection results are filtered afterwards

3.3. Traffic Graph

In addition to the previous data, we also provide a sample traffic graph detailing the location of each camera, as well as the adjacency and distance matrix. The adjacency matrix and distance matrix are created by connecting the cameras that run along a straight road, with some of them located at the intersection between 2 or more roads, or located at a roundabout. In addition, the graph itself contains both directed and undirected sections, where the former represents one-way roads and the latter for two-way roads. The cameras themselves are only connected if they run along the same route (with some sections having 2 different names, despite being on the same route), and have at least one lane directed from one camera to another. Since we are limited to considering the roads that have 3 or more cameras, there is still plenty of room for improvement in subsequent studies, where a denser graph may be used. Also, the sample graph provided is directional, provided that some road sections are one-way only. There are a total of 48 road sections that are sampled, 14 of which are one-way street sections. Originally, there were 140 cameras included in the dataset. However, for simplicity, we have identified 19 clusters of cameras, which are essentially areas where the distances between the cameras are close to each other, such as at crossroads and roundabouts. As for their distance, we determine the distance from one cluster to another to be the distance between the 2 centroids, and similarly, the distance from one camera to a cluster is the same as the distance from that particular camera to the centroid of the cluster. In total, there are 115 separate nodes that are currently being used in this sample graph. Note that this is only our sampled graph, and any subsequent studies are free to construct their own to better fit their needs. As such, we have also provided a fully connected distance matrix, along with the original configuration of the cameras without being clustered. For the distance matrix, we relied on the distance between 2 nodes as reported by Goong.io using their DistanceMatrix API, with the option of using the car as the transportation mode.

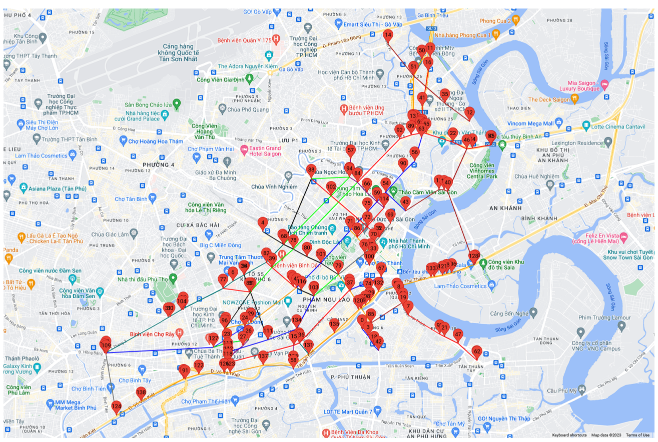


Fig. 2: Distribution of cameras within the dataset across Ho Chi Minh City

4. BENCHMARK TASK

4.1. Experiment setup

For our benchmark task, we experimented with our dataset using 7 baseline methods: historical average, linear regression, support vector regression, random forest regressor, and a simple artificial neural network with 1 hidden layer. The experiment code is publicly available on [GitHub] [4], based on the STREETS dataset benchmark [3] However, unlike their original experiment, which predicts traffic state, traffic count, and vehicle counting, we focus solely on traffic count, aligning with Ho Chi Minh City's traffic characteristics. Additionally, we utilized the adjacency matrix, leaving the distance matrix for future research.

To preprocess the data for modeling, we combined separate vehicle counts into a single traffic count, accounting for variations in vehicle types across timestamps. We then loaded neighborhood data for each sensor, including counts from neighbors within K hops, considering potential errors from faulty sensors. Next, we interpolated the data using nearest-neighbor interpolation to ensure consistent daily sample counts throughout the dataset. Although this method helps maintain consistency, occasional data collection issues such as delayed starts or early endings may occur. Nonetheless, this approach minimizes irregular intervals during data collection to a certain extent.

For model training, we utilized scikit-learn implementations for RFR, SVR, and Linear Regression, and PyTorch for designing the ANN with 2 hidden layers. Unlike the STREETS authors, we don't separate model outputs by traffic states; instead, we predict the vehicle count directly for future timestamps. Evaluation is based on Mean Absolute Error (MAE), comparing model predictions ($\hat{y}_i$) to actual traffic counts (y_i) for N samples. The MAE score is calculated as:

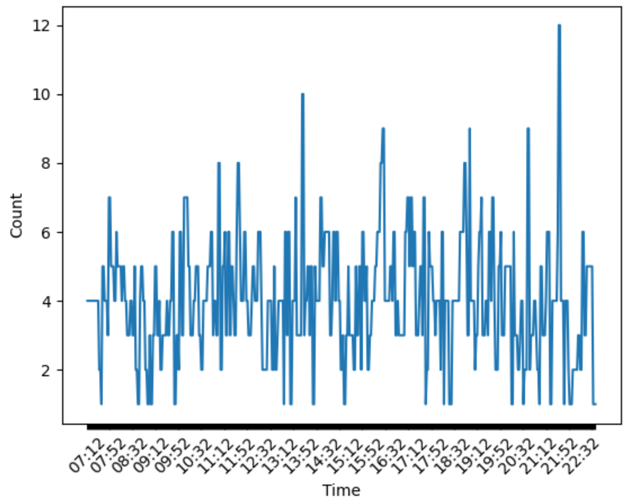
$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{1}$$

In this experiment, we use approximately 1 month of training data from April 5th to May 4th, 2022, and conduct testing on dates between May 5th and May 9th, excluding anomalies on May 5th and May 7th due to crawler restarts. We interpolate at 5-minute and 2-minute intervals instead of solely using 5 minutes, considering Ho Chi Minh City's traffic, which mainly comprises motorcycles. Experiment parameters allow models up to 6 timestamps in the past and predict up to 9 timestamps in the future. Spatial data across nodes includes traffic counts of up to 2 neighbors. All other hyperparameters align with the STREETS GitHub repository. This demonstrates the variability in transportation conditions across sensors and time in Ho Chi Minh City, as illustrated in the figure below.

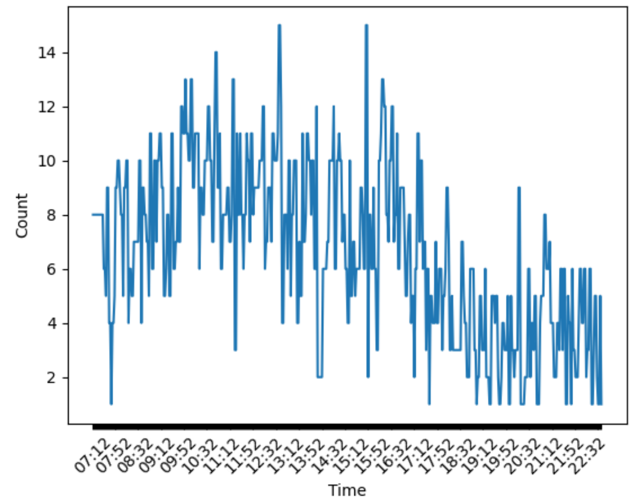
4.2. Experiment results

The results of this experiment are shown in Figures 4 and 5. Figure 4 shows the results when running the experiment on sensor 0, while Figures Figure 5 is obtained from sensor 97, both are run with a 2-minute interval between data points. Sensor 0 is located in a less frequent traffic area (Hoang Dieu

[4]https://github.com/peternguyen39/hcmc_traffic_benchmark



(a) Traffic count of sensor 0 on Hoang Dieu - Khanh Hoi intersection on May 6th 2022, interpolated to every 2 minutes



(b) Traffic count of sensor 97 on Ba Thang Hai - Ly Thuong Kiet intersection on May 6th 2022, interpolated to every 2 minutes

- Khanh Hoi intersection located in District 4), compared to sensor 97 (Ba Thang Hai - Ly Thuong Kiet intersection in District 10), and is relatively more stable, while sensor 97 more accurately reflects the traffic that corresponds to the rush hours and daily commute of Ho Chi Minh City residents. Due to the aforementioned characteristics of traffic in Ho Chi Minh City, we are reporting the results obtained when using 2-minute intervals, allowing the models to see 3 and 6 previous timestamps respectively while forcing the models to predict at 3 timestamps in the future. In addition, we are only selecting the best results within 10 consecutive runs for each method, with each configuration. The best result for each day is highlighted in bold for each of the two tested sensors.

Among the baseline methods applied to sensor 0's data, the artificial neural network (ANN) provides the closest results to the ground truth, across all days. However, for sensor 97, the story is quite different, as no single method outperforms all others in a consistent manner. This can be the result of the data of sensor 0 being relatively more stable throughout the day, as opposed to the large fluctuation between daytime and nighttime as found in the more central area where sensor 97 is located. In addition, the ANN can more closely match the little fluctuations and short-term trends of the data, when roughly compared to other methods.

In addition, since the data collection method is subject to multiple external factors (such as Internet connection bandwidth, and crawler's availability, among other things), the dataset itself is still very limited and hard to predict, even with our best effort. With all of that being said, we firmly believe that this dataset can be useful and indicative of what the actual traffic conditions are in developing countries with densely populated urban areas, not just in Vietnam. There are still various conditions our dataset can provide that we have barely scratched the surface with our experiment.

model	K	history	2022-05-05	2022-05-06	2022-05-07	2022-05-08	2022-05-09
ha	0	3	1.398	1.432	1.519	1.441	1.363
	0	6	1.394	1.435	1.517	1.437	1.360
	1	3	1.398	1.432	1.519	1.441	1.363
	1	6	1.394	1.435	1.517	1.437	1.360
rfr	0	3	1.325	1.566	1.533	1.455	1.411
	0	6	1.242	1.541	1.576	1.451	1.406
	1	3	1.419	1.500	1.517	1.525	1.446
	1	6	1.458	1.523	1.664	1.433	1.424
svr	0	3	1.395	1.418	1.484	1.355	1.341
	0	6	1.395	1.433	1.490	1.355	1.343
	1	3	1.443	1.417	1.500	1.342	1.349
	1	6	1.439	1.430	1.494	1.344	1.352
ann	0	3	1.322	**1.342**	**1.364**	1.430	1.310
	0	6	1.333	1.352	1.373	1.420	1.320
	1	3	**1.149**	1.385	1.366	1.354	**1.292**
	1	6	1.305	1.363	1.397	**1.332**	1.318
linear	0	3	1.254	1.417	1.374	1.446	1.332
	0	6	1.265	1.438	1.421	1.434	1.358
	1	3	1.262	1.415	1.377	1.439	1.330
	1	6	1.269	1.436	1.423	1.431	1.357
linear-lasso	0	3	1.398	1.368	1.395	1.427	1.345
	0	6	1.398	1.375	1.396	1.418	1.345
	1	3	1.398	1.368	1.395	1.427	1.345
	1	6	1.398	1.375	1.396	1.418	1.345
linear-ridge	0	3	1.254	1.417	1.374	1.446	1.332
	0	6	1.265	1.438	1.421	1.434	1.358
	1	3	1.262	1.415	1.377	1.439	1.330
	1	6	1.269	1.436	1.423	1.431	1.357

Fig. 4: Baseline model results for sensor 0 using 2-minute intervals, with K denoting the maximum number of hops for the neighboring traffic sensors

5. DISCUSSION

Our experiments so far have shown that this dataset still has plenty of room for future work, and can be utilized as a means to investigate and draw insights that are previously unavailable or unexplored within the field of traffic prediction and analysis. Although our experiment is still, in more ways than one, limited compared to what could have been done, it provides a useful insight that, to the best of our abilities, determined to only be the tip of the iceberg, both in terms of its potential and its applicability in real-world scenarios.

model	K	history	2022-05-05	2022-05-06	2022-05-07	2022-05-08	2022-05-09
ha	0	3	1.904	2.066	1.753	2.030	1.926
	0	6	1.913	2.062	1.761	2.033	1.922
	1	3	1.904	2.066	1.753	2.030	1.926
	1	6	1.913	2.062	1.761	2.033	1.922
rfr	0	3	1.815	2.194	1.965	2.212	1.921
	0	6	1.920	2.184	1.953	2.210	2.006
	1	3	2.255	**2.005**	1.834	2.071	1.892
	1	6	2.013	2.091	1.777	2.142	1.951
svr	0	3	1.835	2.045	1.762	2.027	1.882
	0	6	1.872	2.044	1.781	2.037	1.879
	1	3	2.132	2.035	1.828	**1.991**	1.908
	1	6	2.168	2.069	**1.752**	2.003	1.945
ann	0	3	1.872	2.656	2.167	2.212	**1.844**
	0	6	1.878	2.644	2.171	2.231	2.606
	1	3	1.789	2.074	1.806	1.996	1.946
	1	6	1.932	2.105	1.771	2.030	1.913
linear	0	3	**1.730**	2.162	1.973	2.211	1.981
	0	6	1.815	2.169	1.988	2.235	2.000
	1	3	1.873	2.174	1.945	2.123	1.974
	1	6	1.937	2.181	1.909	2.173	2.026
linear-lasso	0	3	2.113	2.224	2.082	2.161	2.139
	0	6	2.120	2.229	2.083	2.171	2.145
	1	3	2.113	2.224	2.082	2.161	2.139
	1	6	2.120	2.229	2.083	2.171	2.145
linear-ridge	0	3	1.730	2.162	1.973	2.211	1.981
	0	6	1.815	2.169	1.988	2.235	2.000
	1	3	1.873	2.174	1.945	2.123	1.974
	1	6	1.937	2.181	1.909	2.173	2.026

Fig. 5: Baseline results for sensor 97 with 2-minute interval, with K denoting the maximum number of hops for the neighboring traffic sensors

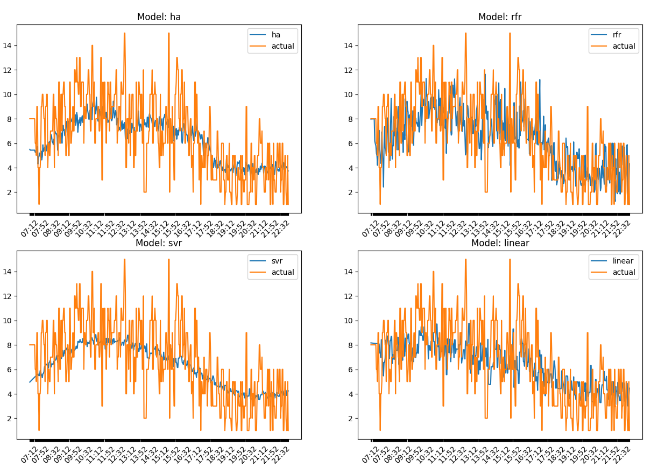


Fig. 6: Results when running 4 different models on sensor 97, at 2 minute interval on May 6th 2022 traffic data

As for future work, we recognized that this dataset can be used for a variety of purposes, including, but not limited to traffic prediction, traffic density estimation, high-density vehicle recognition and under image quality constraints, among others. Since we are releasing the raw image dataset, as well as an instance of the YOLOv5 models that were specifically trained on a subset of said data, there are plenty of room for improvements, both in terms of model architecture, complexity, as well as methodology. For instance, the model that we used - YOLOv5, has seen numerous enhancements, as well as having its eventual successor - YOLOv8 introduced. During the time for this paper, there were also modifications made to the public website at http://giaothong.hochiminhcity.gov.vn/, including refactoring of camera locations and such, as far as we know. Furthermore, since our choice of sensor and other model hyperparameters are following the STREETS authors without conducting any further experiment, the results were, understandably and inevitably, still leave much to be desired.

Furthermore, there is another approach that we believe would better suit the characteristics of our dataset, which is to model the traffic conditions in the area in the form of vehicle coverage of the road area. In other words, we are no longer concerned with the actual number of vehicles on the road, but rather how much percentage of the road section is covered or obstructed by occupants. This method can mitigate the problem of low-resolution cameras, to a certain extent, since it does not need to perform object detection on small, faraway objects such as motorcycles but instead, focus on the general area and shape where the street section is obstructed. This can be done using a number of pre-existing methods, such as background subtraction in a number of studies[18] [19] [20]. Overall, this approach provides a promising alternative for traffic monitoring, in cases where high-resolution cameras are not available, as well as possibly increasing the accuracy of prediction models in cases, such as our dataset.

6. CONCLUSION

Throughout this paper, we have described various aspects of our newly released dataset, as well as performed a relatively simple experiment to prove that there exists a potential that can be further explored by future researchers in the field of urban development, and more specifically, traffic prediction. Our dataset's main contribution lies in its practicality as a representative of the traffic monitoring conditions that are commonly found in developing countries and more specifically, the rapidly growing urban areas. We have also conducted an experiment to evaluate the usability of the dataset, as well as showcased the potential for further research. Even though the dataset itself contains a number of limitations, both in terms of the volume, and the quality of the data, we believe that it would still considerably impact the future development of other researchers working on numerous tasks that can be utilized using our dataset, similarly to how the STREETS dataset have been used as a standard baseline since its release.

7. REFERENCES

[1] Hannah Ritchie and Max Roser, "Urbanization," *Our World in Data*, 2018, https://ourworldindata.org/urbanization.

[2] "Smart cities," Oct 2022.

[3] Corey Snyder and Minh Do, "Streets: A novel camera network dataset for traffic flow," in *Advances in Neural Information Processing Systems 32*, H. Wallach,

H. Larochelle, A. Beygelzimer, F. d'Alché Buc, E. Fox, and R. Garnett, Eds., pp. 10242–10253. Curran Associates, Inc., 2019.

[4] Yaguang Li, Rose Yu, Cyrus Shahabi, and Yan Liu, "Diffusion convolutional recurrent neural network: Data-driven traffic forecasting," in *International Conference on Learning Representations (ICLR '18)*, 2018.

[5] William Chong, J.-H Ng, Srithar Rajoo, and Cheng Tung Chong, "Passenger transportation sector gasoline consumption due to friction in southeast asian countries," *Energy Conversion and Management*, vol. 158, pp. 346–358, 02 2018.

[6] "Roadway sensors," `https://rno-its.piarc.org/en/its-basics-its-technologies-data-and-information/roadway-sensors`, Accessed: 2023-4-11.

[7] Ugo Fiore, Adrian Florea, and Gilberto Lechuga, "An interdisciplinary review of smart vehicular traffic and its applications and challenges," *Journal of Sensor and Actuator Networks*, vol. 8, pp. 13, 02 2019.

[8] Hosagrahar V Jagadish, Johannes Gehrke, Alexandros Labrinidis, Yannis Papakonstantinou, Jignesh M Patel, Raghu Ramakrishnan, and Cyrus Shahabi, "Big data and its technical challenges," *Communications of the ACM*, vol. 57, no. 7, pp. 86–94, 2014.

[9] Binbing Liao, Jingqing Zhang, Chao Wu, Douglas McIlwraith, Tong Chen, Shengwen Yang, Yike Guo, and Fei Wu, "Deep sequence learning with auxiliary information for traffic prediction," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2018, pp. 537–546.

[10] Allister Loder, Lukas Ambühl, Monica Menendez, and Kay W. Axhausen, "Understanding traffic capacity of urban networks," *Scientific Reports*, vol. 9, no. 1, pp. 16283, Nov 2019.

[11] "Utd19 - largest multi-city traffic dataset publicly available," `https://utd19.ethz.ch/`, Accessed: 31 Mar 2023.

[12] Allister Loder, Lukas Ambühl, Monica Menendez, and Kay W. Axhausen, "Utd19. understanding traffic capacity of urban networks," 2020-08, Largest multi-city traffic data set publically available. Please cite https://www.nature.com/articles/s41598-019-51539-5 when using the data set. See https://utd19.ethz.ch/ for more information.

[13] Girish Varma, Anbumani Subramanian, Anoop Namboodiri, Manmohan Chandraker, and CV Jawahar, "Idd: A dataset for exploring problems of autonomous navigation in unconstrained environments," in *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2019, pp. 1743–1751.

[14] Shanghang Zhang, Guanhang Wu, Joao P Costeira, and Jose MF Moura, "Understanding traffic density from large-scale web camera data," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 5898–5907.

[15] Vietnam Government, ," *General statistics for Population and households investigation 2019*, Nov 2019.

[16] Vuong Tran Quang, "Road traffic accident patterns and safety policies suggestions in ho chi minh city," *Journal of Traffic and Transportation Engineering*, vol. 10, pp. 41–48, 2022.

[17] Glenn Jocher, "YOLOv5 by Ultralytics," May 2020.

[18] Belmar Garcia-Garcia, Thierry Bouwmans, and Alberto Jorge Rosales Silva, "Background subtraction in real applications: Challenges, current models and future directions," *Computer Science Review*, vol. 35, pp. 100204, 2020.

[19] S Cheung Sen-Ching and Chandrika Kamath, "Robust techniques for background subtraction in urban traffic video," in *Visual Communications and Image Processing 2004*. SPIE, 2004, vol. 5308, pp. 881–892.

[20] Sen-Ching S Cheung and Chandrika Kamath, "Robust background subtraction with foreground validation for urban traffic video," *EURASIP Journal on Advances in Signal Processing*, vol. 2005, no. 14, pp. 1–11, 2005.